DOCUMENT RESUME

ED 388 685 TM 023 721

AUTHOR Griffin, Bryan W.
TITLE Analysis of Experimental Data via Poisson Regression.
PUB DATE Apr 95
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Analysis of Variance; *Comparative Analysis; *Experiments; *Minimum Competency Testing; *Regression (Statistics)
IDENTIFIERS *Multiple Linear Regression; *Poisson Process

ABSTRACT

This paper provides an exposition on simple and complex comparisons within the framework of Poisson regression. Poisson regression is well suited for the analysis of event count outcomes. Since simple and complex comparisons with both analysis of variance (ANOVA) and multiple linear regression (MLR) are common, the similarities between Poisson regression and ANOVA and MLR are highlighted in example analyses to enhance and clarify the discussion of comparisons within Poisson regression. Although the hypothetical data used for illustration were experimental, the comparison procedures discussed may be applied directly to the analysis of nonexperimental data as well. An appendix provides retake data for the hypothetical minimum competency tests. (Contains 1 figure, 6 tables, and 43 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running head: POISSON REGRESSION

ED 388 685

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

BRYAN W. GRIFFIN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Analysis of Experimental Data Via Poisson Regression

Bryan W. Griffin

Georgia Southern University

Paper presented at the annual meeting of the American Educational Research Association, April 1995, San Francisco, California.

Correspondence concerning this paper should be addressed to Bryan W. Griffin, Department of Educational Leadership, Technology, and Research, Georgia Southern University, Landrum Box 8143, Statesboro, Georgia, 30460-8143. (tele: 912 681 5307)

BEST COPY AVAILABLE

Abstract

This paper provides an exposition on simple and complex comparisons within the framework of Poisson regression. Poisson regression is well suited for the analysis of event count outcomes. Since simple and complex comparisons with both analysis of variance (ANOVA) and multiple linear regression (MLR) is common, the similarities between Poisson regression and ANOVA and MLR are highlighted in example analyses to enhance and clarify the discussion of comparisons within Poisson regression.

Analysis of Experimental Data Via Poisson Regression

Oftentimes variables of interest are not normally distributed. A relatively common non-normal distribution encountered is the Poisson. Examples of data which may be distributed as a Poisson random variate include the number of: (a) tardies recorded for students during a given time frame; (b) changes in programs of study during students' years in college; (c) behavioral problems children display during a school day; and (d) times a student is expelled from school. As these examples illustrate, the Poisson distribution is positively skewed and contains counts (i.e., non-negative integers, such as 0, 1, 2, ..., i) of events. For such counts to follow the Poisson distribution the events must be rare, that is, the expected number of events is very small (e.g., usually less than four).

The Poisson distribution has a distinguishing property: the mean of a Poisson distribution is equal to its variance, i.e., $E(Y) = Var(Y) = \mu$ (Kleinbaum, Kupper, & Muller, 1988). The mean and variance of the Poisson distribution is often denoted by the symbol λ (e.g., Greene, 1990; Lunneborg, 1994). Figure 1 illustrates four randomly generated Poisson distributions. Note the these distributions have means and variances that are approximately equal. Readers wishing a more thorough discussion of the Poisson distribution should see Box, Hunter, and Hunter (1978); Dowdy and Wearden (1991); Kleinbaum, Kupper, and Muller (1988); Neter and Wasserman (1966); or Stockton and Clark (1975).

Modeling outcomes in Poisson regression resembles, in all practical matters, statistical modeling in analysis of variance (ANOVA) and multiple linear regression (MLR). For example, the analysis of simple (e.g., condition A versus B; Keppel & Zedeck, 1989) and complex comparisons (e.g., conditions A and B versus C) in Poisson

regression parallels comparison procedures in MLR. This essay will present an illustrated discussion of simple and complex comparisons in Poisson regression. To facilitate exposition of these procedures, hypothetical data will be used for examples. Since Poisson regression is often directly analogous to ANOVA and MLR, the similarities will be emphasized to enhance the discussion. Readers unfamiliar with concepts in ANOVA and MLR should first consult texts that discuss the two, such as Pedhazur (1982), Cohen and Cohen (1983), or Keppel and Zedeck (1989).

Pairwise Comparisons

Hypothetical data regarding the use of minimum competency tests (MCTs) in high schools is used to illustrate comparison procedures in Poisson regression. The use of MCTs has become an important policy issue due (a) to the requirement in many states that students pass an MCT before a diploma will be granted (Heidorn & Griffin, 1993), and (b) to the supposed negative impact MCTs will have on student behaviors such as dropping out of school before graduation (Archer & Dresden, 1987; Kreitzer, Madaus, & Haney, 1989; LeCompte & Dworkin, 1991; McDill, Natriello, & Pallas, 1987).

Most students pass MCTs (between 60 and 85%) on their first sitting for the examination (Catterall, 1989), and practically all students who remain in school pass the test by the end of their senior year (Serow, 1984). Some states have adopted programs to increase the probability that students will pass the MCT the first time they take it (e.g., Archer & Dresden, 1987), and most states have some remediation program for students who initially fail the MCT. A question of research interest is whether some programs are more successful at reducing the number of times students must retake the MCT before eventually receiving passing marks.

As an example, suppose three small rural high schools are selected for an experiment to assess the impact of two programs on reducing the frequency of MCT retesting before eventual success. The schools are judged to be approximately equal in terms of standardized scores and demographic variables. The two experimental programs are randomly assigned to two of the schools and the third school serves as the control group. The first program, denoted as PRACTICE, provides students with a series of practice examinations. These practice examinations are developed using the same objectives covered on the MCT. The second program, referred to as TUTOR, involves pairing students who are identified as likely to fail the MCT with academically stronger students. The students will spend 30 minutes a day for two weeks working on basic skills assessed by the MCT. Both of these programs focus on preventive measures as opposed to remediation; that is, both are designed to prevent initial failure on the MCT.

As previously noted, the third school will serve as a control and will not implement any program. The experiment is to last for four years and the dependent variable of interest is the number of times students retook the MCT before eventually passing the test or dropping out of school. The primary research questions of interest are: Do the two experimental schools have lower MCT retake rates than the control school, and do either of the two programs appear to be more successful?

Descriptive statistics for the hypothetical data are presented in Table 1. (The raw data is provided in Appendix A in the form of frequencies.) Note that the data appear to follow a Poisson distribution since for each of the three groups, the mean, λ_i (where i = 1, 2, or 3), is approximately equal to the group variance. Also note that the event of multiple retakes of the MCT is relatively rare since most students passed the MCT on their first or second attempt.

The simplest procedure for determining whether statistically significant differences exist in retake rates among the treatments is through pairwise comparisons. Dummy variables are often used in MLR to perform pairwise comparisons (e.g., see Cohen & Cohen, 1983; Keppel & Zedeck, 1989; Pedhazur, 1982), and the usage of dummy variables in Poisson regression is identical to their usage in MLR (Kleinbaum, Kupper, & Muller, 1988; Lunneborg, 1994).

Since there are three categories in the independent variable, type of prevention program, P - 1 = 3 - 1 = 2 dummy variables are needed to represent the P = 3 groups (Pedhazur, 1982). The formulation of the Poisson equation with dummy variables to represent differences in the retake rates between the three groups is analogous to a one-way ANOVA. The ANOVA-like Poisson model with dummy variables is:

$$\ln(\lambda_i) = b_0 + b_1 \text{ PRACTICE} + b_2 \text{ TUTOR} \quad (1)$$

where the right hand side of Equation 1 contains the dummy variables, PRACTICE (coded 1 if student is in school with practice program, 0 otherwise) and TUTOR (coded 1 if student is enrolled in school with the tutor program, 0 otherwise), and the regression coefficients, b_0, b_1, and b_2. Students in the control group represent the omitted or comparison group, and are therefore represented in Equation 1 when both PRACTICE and TUTOR equal 0 (Pedhazur, 1982; Cohen & Cohen, 1983). The expression $\ln(\lambda_i)$ is the natural logarithm of the expected (or mean) number of retakes for the control group. Deviations from the control group mean in natural logarithm units for the PRACTICE and TUTOR groups are represented in the coefficients b_1 and b_2, respectively.

The coefficients b_1 and b_2 in Equation 1 represent additive, linear changes in the natural logarithm of the mean number of retakes

for the groups. For example, a positive b_1 indicates that the log of the mean number of retakes for the PRACTICE group is greater than the log of the mean for the comparison group, CONTROL. Since the natural logarithm preserves the inherent order of the means, a positive b_1 also indicates that the mean number of retakes for the PRACTICE group is greater than the mean number of retakes for the CONTROL group. A negative coefficient is interpreted in a similar fashion, except that a negative b_1 indicates that the PRACTICE mean number of retakes is less than the CONTROL mean. The intercept, b_0, represents the natural logarithm of the mean number of retakes for the CONTROL group. A multiplicative interpretation of the coefficients is presented below shortly.

Why not use ANOVA or MLR?

A Poisson variable cannot take negative values; that is, when counting the number of events that occur, the possible values range, in whole numbers, from 0 to positive infinity. Because of this restriction in the Poisson distribution the use of MLR or ANOVA is inappropriate for several reasons. First, both MLR and ANOVA represent classical normal linear regression models, and parameter estimates for both are obtained by ordinary least squares (OLS). With OLS, one must assume, for proper statistical inference (i.e., point and interval estimation, and hypothesis testing; Gujarati, 1988), that the residuals have a mean of zero, constant variance for all values of the explanatory variables (i.e., homoscedasticity), and are independently distributed (Gujarati, 1988; Wonnacott & Wonnacott, 1979). Due to the nature of the Poisson distribution, the mean and variance will be related, and this ensures non-constant variance, i.e., heteroscedastic (King, 1988). Recall, for example, the data presented in Table 1 and

in Figure 1. Note that as the mean for each distribution increases, the variance also increases.

Another plausible problem with using OLS to obtain parameter estimates for Poisson models is that one may obtain predicted values outside the range of a Poisson variable. It is impossible, for example, for a Poisson variate to take negative values, but estimates obtained from OLS may predict unrealistic negative counts. OLS assumes a linear relationship in the parameters (Gujarati, 1988; King, 1988; Pedhazur, 1982), and such a relationship is not possible with an outcome that follows a Poisson distribution. King (1988) presents a more detailed discussion of the potential problems one might encounter with the use of OLS for modeling Poisson outcomes.

Omnibus Tests

In the general framework of ANOVA one is often interested in determining whether any of the conditions, treatments, or groups have statistically different mean responses on the dependent variable. To determine if statistical differences exist requires an omnibus or overall hypothesis test for the equivalence of the groups, e.g., H_0: $\mu_1 = \mu_2 = \mu_3$ for ANOVA (Keppel, 1991), or H_0: $b_1 = b_2 = 0.00$ for MLR (Kleinbaum & Kupper, 1978). With the current example the omnibus test for the equivalence of retakes rates across the three schools may be expressed as H_0: $b_1 = b_2 = 0.00$.

With ANOVA, an F ratio is used to test H_0. In Poisson regression, omnibus tests of statistical significance are based upon the likelihood ratio test (Kleinbaum, Kupper, & Muller, 1988; Lunneborg, 1994). The likelihood ratio test statistic, or ΔDEV[1], is calculated as the difference in deviances of two nested models (Agresti, 1990; Kleinbaum, 1994). The deviance (DEV) serves as the Poisson regression analog of the residual sums of squares in ANOVA (Aitkin, Anderson,

Francis, & Hinde, 1989), and is defined as -2 x (log of the likelihood, LL), or -2LL (Hosmer & Lemeshow, 1989; Norusis, 1992). Most software usually reports either DEV or the log of the likelihood for Poisson regression.

The likelihood ratio test of the omnibus null, H_0: $b_1 = b_2 = 0.00$, is calculated as follows: $\Delta DEV = DEV_{null} - DEV_{fitted}$, where DEV_{null} is the value of the deviance for the null model (the null model contains only an intercept, e.g., $\ln(\lambda_i) = b_0$), and DEV_{fitted} is the deviance for the fitted model, e.g., Equation 1. The ΔDEV has a chi-square distribution with degrees of freedom (df) equal to the difference in the number of parameters estimated between the two models. That is, df for the likelihood ratio test is determined by the number of parameters estimated in the fitted model minus the number of parameters estimated in the null model.

For the current example (see Table 2 for estimates), the likelihood ratio test is calculated as follows: DEV_{null} is 365.145, and DEV_{fitted} for Equation 1 is 350.876, so $\Delta DEV = 365.145 - 350.876 = 14.269$. Equation 1 contains three terms, b_0, b_1, and b_2; and the null model contains only an intercept, b_0. Therefore, the number of parameters estimated in Equation 1 is 3, and the number estimated in the null model is 1, so $df = 3 - 1 = 2$. The p-value for a ΔDEV of 14.269 with df equal to 2 is .0008, which indicates that the MCT retake rates differ, statistically, across the three schools.

Note that the likelihood ratio test also may be used in a fashion analogous to partial F tests in ANOVA. That is, the likelihood ratio test is well suited for testing differences in fit between hierarchically nested models (Agresti, 1990; Cohen & Cohen, 1983; Collett, 1991). For illustrations, see Lunneborg (1994) or Kleinbaum, Kupper, and Muller (1988).

Multiple Comparisons and Control of the Type 1 Error Rate

Since the omnibus likelihood ratio test of Equation 1 is statistically significant, the next step in the analysis is to determine precisely which of the schools have statistically different MCT retake rates. As in ANOVA and MLR, one may formulate and test pairwise comparisons in Poisson regression. Since there are three schools, there are three possible comparisons: (1) PRACTICE versus CONTROL, (2) TUTOR versus CONTROL, and (3) PRACTICE versus TUTOR. The first and second comparisons are represented directly by b_1 and b_2, respectively, in Equation 1.

Examination of the estimates in Table 2 indicates that the difference between the retake rates for PRACTICE and CONTROL is statistically significant (i.e., $z = b_1/se_{b_1} = -3.24$, $p < .05$)[2]. As noted, the coefficient representing this difference is b_1. This coefficient is negative which indicates that students exposed to the PRACTICE treatment have lower MCT retake rates than students in the CONTROL school. The mean number of retakes for the CONTROL group is $\exp(b_0) = \exp(-0.916) = .400$, (where exp() represents the exponent of the regression coefficient). The mean number of MCT retakes for students exposed to the PRACTICE treatment is $\exp(b_0 + b_1) = \exp(-0.916 + -0.981) = \exp(-1.897) = .150$. The ratio of the number of MCT retakes of the PRACTICE to CONTROL group is .15 / .40 = 0.375. Calculating the exponent of b_1 provides this ratio, that is, $\exp(b_1) = 0.375$. The exponent of b_1 may be interpreted as follows: The mean number of MCT retakes for the PRACTICE group is 0.375 times lower than the mean number of retakes for the CONTROL group.

The difference in MCT retake rates between TUTOR and CONTROL is also statistically significant ($z = -2.77$, $p < .05$). The exponent of b_2 is $\exp(b_2) = 0.45$. Thus, students exposed to the TUTOR treatment

have a mean number of retakes that is 0.45 times lower than the mean number of retakes for the CONTROL group.

Examination of the pairwise comparison between PRACTICE and TUTOR will require one of two approaches: (1) recode the dummy variables so either PRACTICE or TUTOR becomes the comparison group; or (2) manually compute the difference in the log of retake rates between PRACTICE and TUTOR and then divide by the standard error of this difference.

The first method for calculating the PRACTICE and TUTOR comparison calls for a respecification of the dummy variables. Consider, for example, the following model:

$$\ln(\lambda_i) = b_0 + b_1 \text{ PRACTICE} + b_2 \text{ CONTROL}. \qquad (2)$$

As in Equation 1, b_1 is the change in the logrithm of the retake rate for the PRACTICE group, but now the comparison group is TUTOR. The intercept, b_0 is the natural logrithm of the retake rate for students exposed to the TUTOR treatment, and b_2 is the change in the logrithm of the retake rate for the CONTROL students.

The focus in Equation 2 is upon b_1 since this coefficient provides the comparison between PRACTICE and TUTOR. The parameter estimate for b_1 in Equation 2 is 0.192, the standard error is 0.35, and the ratio b_1/se_{b_1} is 0.55. The corresponding p-value is .582, so the difference in mean retakes between PRACTICE and TUTOR is not statistically significant.

Manual computation of the estimated difference between PRACTICE and TUTOR is only slightly more complex. The general formula for the difference between two groups, as represented by dummy variables, is $b_j - b_k$. The standard error of this difference is obtained by

$$se_d = \sqrt{s_{b_j}^2 + s_{b_k}^2 - 2s_{b_j}s_{b_k}}$$

where se_d is the standard error of the difference, $s_{b_j}^2$ is the estimated variance of b_j, $s_{b_k}^2$ is the variance of b_k, and $2s_{b_j}s_{b_k}$ is the covariance between b_j and b_k multiplied by two.

To illustrate, the covariance between b_1 and b_2 (from Equation 1) is 0.025; the standard errors for b_1 and b_2 are 0.303 and 0.283, respectively. With this information, the standard error of the difference between b_1 - b_2 may be calculated as follows:

$$se_d = \sqrt{0.303^2 + 0.283^2 - (2 \times 0.025)}$$

$$= \sqrt{0.122}$$

$$= 0.35.$$

The corresponding ratio is $[(b_1 - b_2) / se_d] = [(-0.981 - (-0.788)) / 0.35] = -0.55$. This calculated z ratio is less than the .05 level critical z value ($z_{crit} = \pm 1.96$), therefore the null, H_0: $b_1 - b_2 = 0.00$, is not rejected. As a result, one may conclude that students exposed to the PRACTICE and TUTOR treatments appear to have approximately equal retake rates.

The procedure of pairwise comparisons just illustrated followed two steps. First, the statistical significance of the overall null (H_0: $b_1 = b_2 = 0.00$) was assessed using an omnibus likelihood ratio test. Second, following a decision to reject the overall H_0, pairwise comparisons were performed. Taken together, these two steps are identical to Fisher's protected t procedure for multiple comparisons in ANOVA and MLR (Cohen & Cohen, 1983; Maxwell & Delaney, 1990). The problem with this procedure is that as the number of comparisons increases, the probability of making a Type I error also increases;

thus, protected tests fail to control for inflation of the Type I error rate (Maxwell & Delaney, 1990; Toothaker, 1991; but see Cohen & Cohen, 1983; Keppel & Zedeck, 1989).

An alternative that provides better control of Type I errors in Poisson regression is the Bonferroni multiple comparison procedure (DeMaris, 1992; Larntz, 1993; Tarone, 1990; Westfall & Young, 1989). The Bonferroni procedure simply requires that the researcher determine the number of comparisons to be made, denoted C, and then divide nominal α by C, i.e., α / C. The resulting α level then becomes the one used in hypothesis tests of the comparisons. For example, there are three comparisons of interest in the MCT retake data. If the nominal α is set at .05, then the Bonferroni adjusted α is .05 / 3 = 0.0167. If p-values provided by statistical software are used, one simply compares the obtained p-values against the adjusted α level of .0167 to determine statistical significance (Maxwell & Delaney, 1990). For instance, of the three comparisons calculated above, two have p-values smaller than the adjusted α of .0167. The p-value for the PRACTICE versus CONTROL comparison, b_1 from Equation 1, was less than .0167; therefore, the statistical evidence indicates that PRACTICE provides a lower mean number retakes relative to the CONTROL group. In addition, the difference in retake rates between TUTOR and CONTROL is also statistically significant, $b_2 = -0.788$, $p < .0167$. The difference in mean number of retakes between PRACTICE and TUTOR, however, is not statistically significant at the adjusted .0167 alpha level (see Table 2).

If p-values are not accessible, then one may wish to use the corresponding critical z value for the Bonferroni adjusted α. Table 3 provides the appropriate critical z values for nominal αs of .10, .05 and .01 adjusted for up to eight comparisons. For example, in the MCT retake study there are three pairwise comparisons. The Bonferroni

adjusted critical z ratio for an α of .05 and three comparisons is 2.394. The calculated z ratios for the comparisons are computed as was done earlier using b_1 and b_2 and the difference between $b_1 - b_2$. Recall that the absolute value of the calculated z ratio for the PRACTICE versus CONTROL comparison was 3.24 which is larger than the Bonferroni adjusted critical of 2.394, and the absolute value of the calculated z ratio for TUTOR versus CONTROL was 2.78 which is also greater than the critical z of 2.394. The PRACTICE versus TUTOR comparison had an absclute value z ratio of 0.55 which is not larger than the BOnferroni adjusted critical z. Note that inferential results based upon Bonferroni adjusted αs or Bonferroni adjusted critical z values will be identical. Also, if one uses the Bonferroni procedure, then an omnibus likelihood ratio test is not needed for planned comparisons (Maxwell & Delaney, 1990).

Complex Comparisons

It is possible in Poisson regression to construct complex comparisons. The procedure is identical to that described by Pedhazur (1982) and Cohen and Cohen (1983) for MLR. For example, suppose one hypothesized that MCT retake rates will not differ between students exposed to either of the two treatments, PRACTICE and TUTOR; but that 0the retake rates will differ between the CONTROL school and the PRACTICE and TUTOR schools.

Coding of Contrasts

The above compound hypothesis represents two comparisons. The first compares the retake rates between PRACTICE and TUTOR. The second calls for a comparison of the retake rate for CONTROL and the <u>average</u> retake rate for PRACTICE and TUTOR. A standard method of representing these two contrasts, which are referred to as contrasts A and B, respectively, is presented in Table 4. As a concrete example, Table 5

illustrates, for a sample of nine cases, the organization of the data as presented in spreadsheet fashion. Pedhazur (1982, p. 308) provides a similar example for MLR.

The numbers given in Table 4 are the coefficients used to represent the cotrasts in the Poisson regression equation. The corresponding equation is:

$$\ln(\lambda_i) = b_0 + b_1 A + b_2 B \qquad (3)$$

where b_1 is the effect for contrast A, which may be interpreted as the difference in the log of the mean number of retakes between PRACTICE and TUTOR; and b_2 represents the effect associated with contrast B. Contrast B is defined as the difference in the logrithm of the mean number of retakes between CONTROL and the average logrithm mean number of retakes for PRACTICE and TUTOR, i.e., (PRACTICE + TUTOR) / 2.

The coding scheme presented in Table 4 must be considered when interpreting the estimated coefficients. For instance, in contrast B, the CONTROL group is coded with the larger number (2 versus -1). Recall that the dependent variable ranges from 0 to, theoretically, infinity. If the CONTROL group is expected to have a higher retake rate, then b_2 should be positive since a direct relationship would result between the coding of contrast B and the dependent variable. If, however, CONTROL students were expected to have a lower retake rate, then b_2 would be negative.

Coefficient estimates for Equation 3 are provided in Table 6. Contrast A is not statistically significant, and this indicates that the retake rates do not statistically differ between the two treatments. Contrast B, however, is statistically significant. This shows that students exposed to the PRACTICE and TUTOR treatments have retake means that are lower than students in the CONTROL school.

Omnibus Tests and Multiple Comparisons

The omnibus likelihood ratio test of models with complex comparisons is identical to the omnibus test for pairwise comparisons. That is, one examines the change in the deviances between the null and fitted model and compares this change to a chi-square distribution with *df* equal to the difference in the number of parameters in the two models. For example, $\Delta DEV = DEV_{null} - DEV_{full} = 365.145 - 350.876 = 14.269$, with $df = 3 - 1 = 2$ ($p < .01$).

Some behavioral statisticians argue that one need not be concerned with an increase in the Type 1 error rate when a few planned, orthogonal comparisons are performed in ANOVA or MLR (Keppel & Zedeck, 1989; Pedhazur, 1982), but others disagree (Maxwell & Delaney, 1990). The point, however, is of little importance here because orthogonality is not possible in Poisson regression (see Collett, 1991, pp. 68-69, for a similar discussion in logistic regression). Therefore, the Bonferroni adjustments discussed earlier are recommended when examining several comparisons. For example, since there are only two contrasts in the MCT retake data, the Bonferroni adjusted α is .05 / 2 = .025. As before, contrast A is not statistically significant because its *p*-value is larger than the adjusted α level, but contrast B remains statistically significant.

If *p*-values are not provided, Bonferroni adjusted critical *z* values may be used. Referring again to Table 3, the adjusted critical *z*, for two comparisons, is ± 2.241. The calculated *z* (in absolute value) associated with contrast A is 0.55, which is smaller than the critical *z*, and the calculated *z* for contrast B is 3.74, which is larger than 2.241. These inferential results parallel the results obtained from the Bonferroni adjusted α.

The Bonferroni multiple comparison procedure is appropriate as long as the number of comparisons tested remain relatively small. As the number of increases, the Bonferroni procedure tends to lose power relative to other multiple comparison procedures (Maxwell & Delaney, 1990). Rosenthal and Rubin (1984) offer an alternative Bonferroni procedure that addresses the issue of power loss with many comparisons, and for very large numbers of comparisons in Poisson models, more powerful multiple comparison procedures are available (see, for instance, Tarone, 1990; Westfall & Young, 1989).

Summary

The procedures for simple and complex comparisons in Poisson regression are comparable to those in ANOVA and MLR. As with ANOVA and MLR, these procedures may be easily extended to include Poisson models with multiple categorical and continuous variables, and models with interactions. In fact, due to the similarity between Poisson regression and MLR, many of the procedures discussed in standard MLR texts relate directly to Poisson regression. Although the hypothetical data used for illustration above were experimental, the comparison procedures discussed may be applied directly to the analysis of nonexperimental data as well.

Footnotes

[1]The likelihood ratio test statistic is often represented by G^2 (Larntz, 1993), G (Hosmer & Lemeshow, 1989), L^2 (Austin et al., 1992), ΔC^2 (DeMaris, 1992), and H^2 (Jobson, 1992). The symbol ΔDEV was chosen here since it conveys the nature of the likelihood ratio test, i.e., a change, Δ, in the deviance, DEV.

[2]The ratio of the coefficient to its standard error in Poisson regression follows, in large samples, a z distribution. An alternative, and more powerful test, however, is the likelihood ratio.

References

Agresti, A. (1990). Categorical data analysis. New York: Wiley.

Aitkin, M., Anderson, D., Francis, B., & Hinde, J. (1989). Statistical Modelling in GLIM. New York: Oxford.

Aldrich, J., & Nelson, F. (1984). Linear probability, logit, and probit models. Newbury PRACTICE, CA: Sage.

Archer, E. L., & Dresden, J. H. (1987). A new kind of dropout: The effect of minimum competency testing on high school graduatiopn in Texas. Education and Urban Society, 19, 269-279.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. In J.C. Smart (ed.), Higher education: Handbook of theory and research. (pp. 379-410). New York: Agathon.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for experimenters: An introduction to design, data analysis, and model building. New York: Wiley.

Catterall, J. S. (1989). Standards and school dropouts: A national study of tests required for high school graduation. American Journal of Education, 98, 1-34.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Collett, D. (1991). Modelling binary data. London: Chapman & Hall.

DeMaris, A. (1992). Logit modeling: Practical applications. Newbury PRACTICE, CA: Sage.

DeMaris, A. (1993). Odds versus probabilities in logit equations: A reply to Roncek. Social Forces, 71, 1057-1065.

Dowdy, S., & Wearden, S. (1991). Statistics for research. New York: Wiley.

Greene, W. H. (1990). Econometric analysis (2nd ed.). New York: Macmillan.

Griffin, B. W. (1994, April). Analysis of experimental data with logistic regression. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Gujarati, D. N. (1988). Basic econometrics (2nd ed.). New York: McGraw-Hill.

Heidorn, M. H., & Griffin, B. W. (1993). Minimum competency testing and high school dropouts. Paper presented at the meeting of the American Educational Research Association, Atlanta, GA.

Hosmer, D., & Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

Jobson, J. D. (1992). Applied multivariate data analysis (Vol. 2). New York: Springer-Verlag

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Keppel, G., & Zedeck, S. (1989). Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches. New York: Freeman.

Kleinbaum, D. G. (1994). Logistic regression: A self-learning text. New York: Springer-Verlag.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Applied regression analysis and other multivariable methods (2nd ed.). North Scituate, MA: Duxbury.

Knoke, D., & Bohrnstedt, G. W. (1993). Statistics for social data analysis. Itasca, IL: Peacock.

Kreitzer, A. E., Madaus, G. F., & Haney, W. (1989). Competency testing and dropouts. In L. Weis, E. Farrar, & H. G. Petrie (Eds.), Dropouts from school: Issues, dilemmas, and solutions (pp. 129-152). Albany: State University of New York Press.

Larntz, K. (1993). Analysis of categorical response variables. In L. Edwards (ed.), Applied analysis of variance in behavioral science. (pp. 497-540). New York: Marcel Dekker.

LeCompte, M. D., & Dworkin, A. G. (1991). Giving up on school: Student dropouts and teacher burnouts. Newbury Park, CA: Corwin Press, Inc.

Lunneborg, C. E. (1994). Modeling experimental and observational data. Belmont, CA: Duxbury.

Maxwell, S. E., & Delaney, H. D. (1990). Designing experiments and analyzing data: A model comparison perspective. Pacific Grove, CA: Brooks/Cole.

McDill, E. L., Natriello, G., & Pallas, A, M. (1987). A population at risk: Potential consequences of tougher school standards for student dropouts. In G. Natriello (Ed.), School dropouts: Patterns and policies (pp. 106-147). New York: Teachers College Press.

Neter, J., & Wasserman, W. (1966). Fundamental statistics for business and economics (3rd ed.). Boston: Allyn and Bacon.

Norusis, M. (1992). SPSS: SPSS/PC+ advanced statistics version 5.0. Chicago: SPSS Inc.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart, and Winston.

Retherford, R. D., & Choe, M. K. (1992). Statistical models for causal analysis. New York: Wiley.

Roncek, D. W. (1991). Using logit coefficients to obtain the effects of independent variables on changes in probabilities. Social Forces, 70, 509-518.

Roncek, D. W. (1993). When will they ever learn that first derivatives identify the effects of continuous independent variables or "officer, you can't give me a ticket, I wasn't speeding for an entire hour." Social Forces, 71, 1067-1078.

Rosenthal, R., & Rubin, D. B. (1984). Multiple contrasts and ordered Bonferroni procedures. Journal of Educational Psychology, 76, 1028-1034.

Serow, R.C. (1984). Current trends in States' Graduation Policies. Mimeographed, North Carolina State University, Department of Education.

Stockton, J. R., & Clark, C. T. (1975). Introduction to business and economic statistics (5th ed.). West Chicago, IL: South-Western.

Tarone, R. (1990). A modified Bonferroni method for discrete data. Biometrics, 46, 515-522.

Toothaker, L. E. (1991). Multiple comparisons for researchers. Newbury Park, CA: Sage.

Westfall, P., & Young, S. (1989). p value adjustments for multiple tests in multivariate binomial models. Journal of the American Statistical Association, 84, 780-786.

Wonnacott, R. J., & Wonnacott, T. H. (1979). Econometrics (2nd ed.). New York: Wiley.

Wonnacott, T. H., & Wonnacott, R. J. (1981). Regression: A second course in statistics. New York: Wiley.

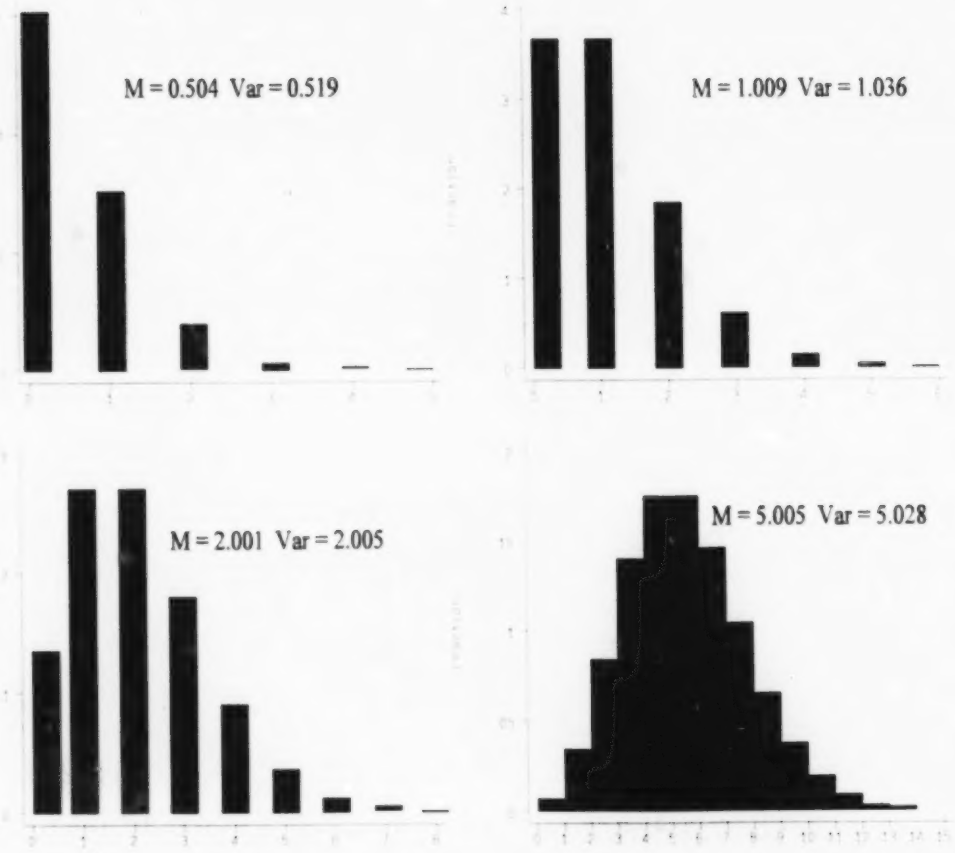
Appendix A

MCT Retake Data

Count of Retakes	Frequency of Retakes		
	School 1: PRACTICE Condition	School 2: TUTOR Condition	School 3: CONTROL School
0	86	82	67
1	13	16	27
2	1	1	5
3			1

Note. The column labeled "Count of Retakes" indicates the number of times a student retook the MCT. For example, a count of 0 signifies no retakes (student took the exam once); a count of 1 indicates one retake of the MCT, etc.

Figure 1



BEST COPY AVAILABLE

Table 1

Descriptive Statistics for the Minimum Competency Test Retake Data

Group	Mean	Variance	Minimum Value	Maximum Value	n
PRACTICE	0.150	0.149	0	2	100
TUTOR	0.182	0.171	0	2	99
CONTROL	0.400	0.404	0	3	100

Table 2

Poisson Regression Estimates for Equation 1

Variables		b	se	b/se	p-value
PRACTICE,	b_1	-0.981	0.303	-3.24	.001
TUTOR,	b_2	-0.788	0.284	-2.77	.005
Intercept,	b_0	-0.916	0.158	-5.80	.000

Note. DEV_{null} = 365.145; DEV_{fitted} = 350.876; ΔDEV = 14.270, df = 2, p-value < .01.

Table 3

Bonferroni Adjusted Critical z Values for Multiple Comparisons

α	Number of Comparisons							
	1	2	3	4	5	6	7	8
.10	1.645	1.960	2.128	2.241	2.326	2.394	2.450	2.498
.05	1.960	2.241	2.394	2.498	2.576	2.638	2.690	2.734
.01	2.576	2.807	2.935	3.023	3.090	3.144	3.189	3.227

Note. The entries in this table were computed by the author.

Table 4

Contrasts for (A) PRACTICE vs. TUTOR

and (B) CONTROL vs. (PRACTICE + TUTOR) / 2

	Experimental Treatment		
Contrast	PRACTICE	TUTOR	CONTROL
A	+1	-1	0
B	-1	-1	+2

Table 5

Example Data for Complex Comparisons

Student Identification Number	Number of MCT Retakes	Contrast A	Contrast B	(Treatment)
1	1	1	-1	(PRACTICE)
2	1	1	-1	(PRACTICE)
3	0	1	-1	(PRACTICE)
4	0	-1	-1	(TUTOR)
5	1	-1	-1	(TUTOR)
6	1	-1	-1	(TUTOR)
7	1	0	2	(CONTROL)
8	2	0	2	(CONTROL)
9	0	0	2	(CONTROL)

Table 6

Poisson Regression Estimates for Equation 3

	b	se	b/se	p-value
Contrast A	-0.096	0.175	-0.55	.582
Contrast B	0.295	0.079	3.73	.000
Intercept	-1.506	0.128	-11.77	.000

Note. DEV_{null} = 365.145; DEV_{fitted} = 350.876; ΔDEV = 14.270, df = 2, p-value < .01.